N68-15834

THE EFFECT OF CHANNEL ERRORS ON DATA COMPRESSION*

Lee D. Davisson
School of Engineering and Applied Science
Princeton University
Princeton, New Jersey

SUMMARY

This paper considers the effects of channel noise on received reconstructed data which is compressed prior to transmission. The energy compression ratio is defined and used as a figure of merit for a given system. Calculations are illustrated for run length coding with periodic synchronization. Noise is shown to have a critical effect on system performance. In fact, if the signal-to-noise ratio is too low, run length coding is found to be worse than no data compression at all from a data quality point of view. Properly employed, an improvement can be gained by using error correcting codes.

INTRODUCTION

There has been much interest in reducing the required bandwidth or equivalently the transmission rate for a given data source whose probability distributions are unknown by taking advantage of any data redundancies which may exist. Methods to do this are called data compression methods. Numerous investigators have theoretically and empirically determined algorithms for data compression. For sampled quantized data the figure of merit for a given system is usually quoted as the compression ratio, the ratio of the transmission rate for the original data using some standard coding method to that of the compressed data in either samples generated per sample transmitted or bits uncompressed per bit compressed so that a high compression ratio is associated with a "good" system. In most previous work the fact that the compressed data must be transmitted over a "real" channel which is noisy is glossed over or ignored - the channel is assumed noiseless. The purpose of this paper is to show theoretically that that is a poor assumption. Even for very low channel error rates the effect on the received reconstructed sample error rate is significant. Thus the compression ratio as defined above is not complete. For a noisy channel <u>the appropriate figure of merit[1] is the required energy per sample</u> to transmit the data for a fixed level of "quality" (where quality is defined by some cost function such as probability of error or mean square error) when the <u>entire</u> system is considered - transmitter, encoder, channel, decoder and receiver systems. The ratio of the energy per sample for some standard system (e.g., PCM) to that in a data compression system for the same level of quality is called the <u>energy</u> compression ratio for that system while the ratio of transmission rates is called the data compression ratio in samples per sample or the bit compression ratio in bits per bit. The calculation of the latter quantities is simple relative to the calculation of the energy compression ratio except for the most trivial systems. Raga[2,3] obtained approximate theoretical and empirical results using a mean square quality criterion for video data with perfect line synchronization, although energy comparisons were not made and the data mathematical model was not complete. Davisson[1] obtained exact theoretical results for a first order Markov process (the approximate case for video data) assuming perfect line synchronization.

In this paper the results of reference [1] are reviewed and extended to indicate the effects of synchronization and error correcting codes. The specific system chosen for illustration is the commonly suggested run length coding wherein the data is encoded as a sequence of level-run length word pairs. The run length word gives the number of times the level is to be repeated. It is assumed that the data is transmitted as a sequence of ones and zeros using coherent detection. Probability of error is chosen as the quality criterion. It is shown that the ratio of the energy required to transmit the samples directly to that required to send the data compressed for a given level of received data quality is considerably less than

*This work was supported under NASA Grant NGR-31-001-068 and through the NASA Goddard Spaceflight Center Summer Workshop Program.

6/2

HARDCOPY $3.00 pages 5
MICROFICHE .65 code 1
Cat 07 OR# 92513

would be predicted from the "noiseless" solution even for error rates as small as 10^{-7} bits/bit. It is demonstrated that when error correcting codes are used properly an improvement can be made. However, if not used well, error correcting codes result in a lower energy compression ratio.

PROBLEM DESCRIPTION

To illustrate the ideas introduced in the preceding section, the energy compression ratio is calculated for a particular coding scheme and a particular data and channel probability model. In detail:

1. The data is encoded in blocks of N samples (e.g., a TV line) preceded by a synchronization word which is never missed. It can be shown that the latter assumption has little effect on the end result.

2. Within each block the data is run length encoded as a level-run word pair. The level word is one of L levels and the run length, the number of times the sample is repeated, is one of T values. Each value is PCM encoded. In some cases an error correcting code is used on the timing word or both words. Although the number of samples N in each block is fixed, the number of code words in each block varies depending on the data. Note that if an error is made in the timing word its effect is cumulative. All succeeding data up to the next synch word is shifted by the amount of the error. Errors are additive and independent so that the shift variance increases linearly with distance from synchronization. Thus from this point of view, synch words should appear frequently. On the other hand, the insertion of a synch word in the data causes the last run in the block to be truncated resulting in two words in place of one (the last word in the present block and the first word of the next) in addition to the synch word itself or, in other words, a reduction in the data compression ratio. Thus there is a tradeoff between timing error propagation and run truncation. This effect is studied in more detail later.

3. The data is a first order Markov chain with a probability p called the probability of prediction, of remaining at the same level independent of the level (which approximates the situation for TV data). The other transition probabilities can be chosen arbitrarily.

4. The channel is binary symmetric. Specifically the bits are coherently transmitted in white Gaussian noise so that if λ^2 is the signal energy-to-noise spectral density ratio, the bit error probability is:

$$\Phi(-\lambda),$$

where $\Phi(\cdot)$ is the Gaussian cumulative distribution function.

5. The data quality measure is the probability of error averaged over the block length (due to the timing error propagation effect errors are more likely at the end of the block than the beginning).

DATA COMPRESSION RATIO

Let M be the smallest integer less than or equal to $\frac{N}{T}$. Then the data compression ratio is calculated for the above model as

$$C_d(N) = \left[\frac{1-p}{1-p^T} + \frac{1}{N} + \frac{p^T}{1-p^T}\left(\frac{p-p^M}{N} + \right.\right.$$

$$(1-p)\,p^M(M-1)$$

$$\left.\left. - \frac{1}{M}\,\frac{(1-p)}{1-p^T}\,(1-p^M)\right)\right]^{-1}.$$

As $N \to \infty$, $C_d(N+1) > C_d(N)$ with the limiting value:

$$C_d(\infty) = \frac{1-p^T}{1-p} .$$

The decrease in data compression ratio with decreasing block length is a result of end of the block run truncation as previously discussed.

BIT COMPRESSION RATIO

The bit compression ratio is smaller than the data compression ratio due to the coding requirements. It is given by:

$$C_b(N) = C_d(N)\frac{\log_2 L}{\log_2 L + \log_2 T + b_{ec} + \frac{s}{N}}$$

where b_{ec} is the number of error correcting bits and s is the length of the synch word in bits.

ENERGY COMPRESSION RATIO

The energy compression ratio is defined as the ratio of the energy per sample required to send the data in PCM form to that in compressed form for a fixed probability of error, P_e. If the signal energy-to-noise spectral density ratios in each system for this error probability are λ^2_{PCM} and λ^2_c respectively, then the energy compression ratio is

$$C_e(P_e,N) = \frac{\lambda^2_{PCM}}{\lambda^2_c} C_b(N).$$

The probability of error for the PCM system is given by

$$P_e = 1 - \Phi^{\log_2 L}(\lambda_{PCM})$$

where $\Phi(\cdot)$ is the cumulative Gaussian probability distribution or:

$$\lambda_{PCM} = \Phi^{-1}\left[(1-P_e)^{1/\log_2 L}\right]$$

where Φ^{-1} is the inverse of Φ.

The details of the probability of error calculations for the compressed data are extremely involved. The reader can refer to reference [1] for the details. Generally speaking, the method involves the calculation of the probability that h timing word errors have preceded a randomly selected sample times the probability of error given h timing errors. This expression can be evaluated numerically starting with one error and summing until further errors contribute little to the probability of error.

NUMERICAL RESULTS

To illustrate the nature of the energy compression ratio and the effect of timing errors on the received data, the probability of received sample error is calculated numerically. It will be seen that the probability of received sample error is greater in the compressed system than in the uncompressed system for the same bit error rate increasing with N to values many times that of the PCM system. The following system parameters are used as being "typical" (specifically, as being typical of TV data):

L = number of levels = 16
T = maximum run length = 16
p = probability of prediction = .8
s = synch bits = 8

The complete Markov transition probability matrix places all the probability weight on the same level and its neighbors. That is the transition probability from level m to level k is

$$Pr(k|m) = \begin{cases} p & k = m \\ \frac{1-p}{2} & k = m\pm 1 \quad 1<m<T \\ 1-p & k = m-1=T-1 \text{ or } k = m+1=2 \\ 0 & \text{elsewhere.} \end{cases}$$

Figure 1 shows the ratio of the sample error rate for compressed data to that for uncompressed data as a function of N when in each case a bit error rate of 10^{-3} is used. The curve is essentially the same for all error rates such that the probability of more than one bit error occuring in any block of N samples is negligible. The curve is approximately linearly increasing within this range. The increasing nature should cause the energy compression ratio to decrease for sufficiently large N. This is verified in figure 2 where the energy compression ratio appears as a function of N for several bit error rates in the data compression system (in the PCM system the bit error rate varies so that the same sample error rate exists in both systems). The energy compression ratio reaches a maximum at an optimum value of N depending on the error rate. For certain values of N, the compression system is worse than no compression at all.

Figure 3 shows the compression ratio as a function of the bit error rate for a fixed value of N=500 (a typical line length for TV data). Note that even at an error rate of 10^{-7} the energy compression ratio is significantly below the bit compression ratio.

Figure 4 gives the energy compression ratio when a (7,4) error correcting code is used on the timing word only, and on both the level and the timing word at a bit error rate of 10^{-3}. It can be seen that if the complexity of error correcting codes is to be added to the system, that error protection should be added to both the level and timing words. As in figure 2, the energy compression ratio decreases monotonically with N above some optimum value showing once again the importance of making a proper choice of block size.

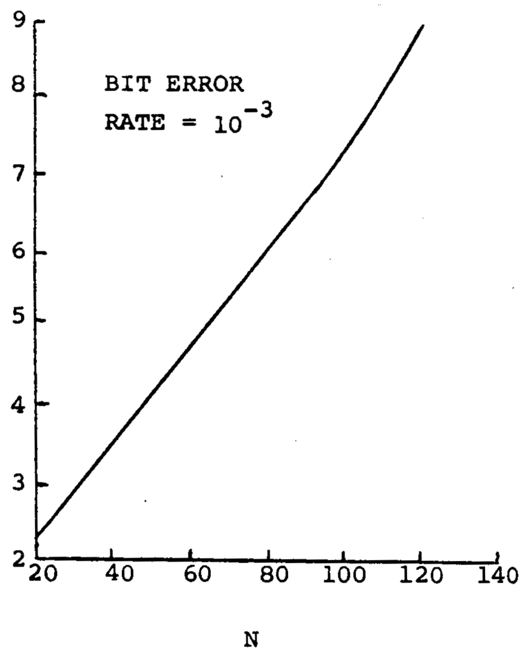

Fig. 1 $\frac{\text{SAMPLE ERROR RATE COMPRESSED}}{\text{SAMPLE ERROR RATE UNCOMPRESSED}}$ vs. N

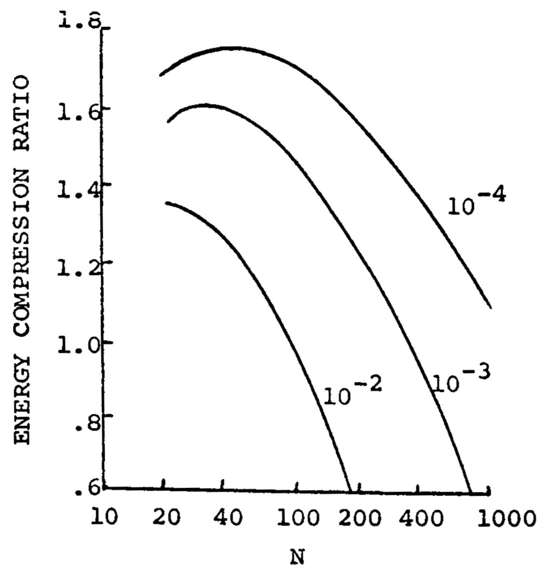

Fig. 2. ENERGY COMPRESSION RATIO vs. N AT BIT ERROR RATES OF 10^{-2}, 10^{-3}, 10^{-4}.

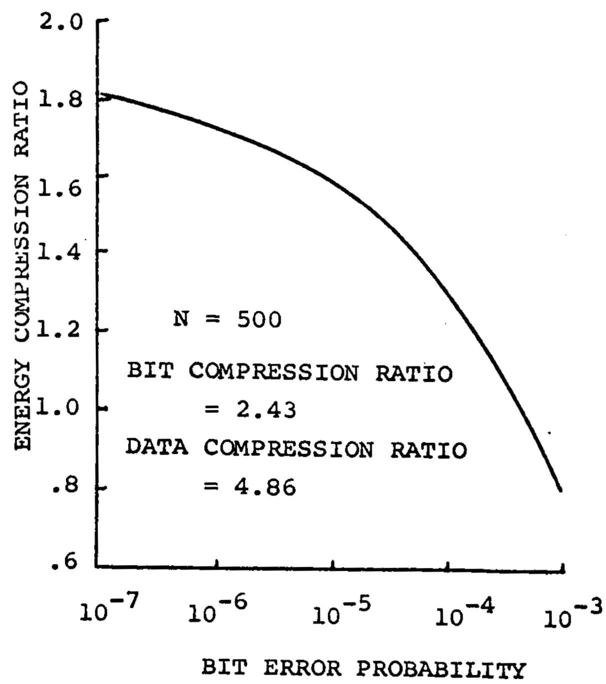

Fig. 3 ENERGY COMPRESSION RATIO vs. BIT ERROR PROBABILITY FOR N=500.

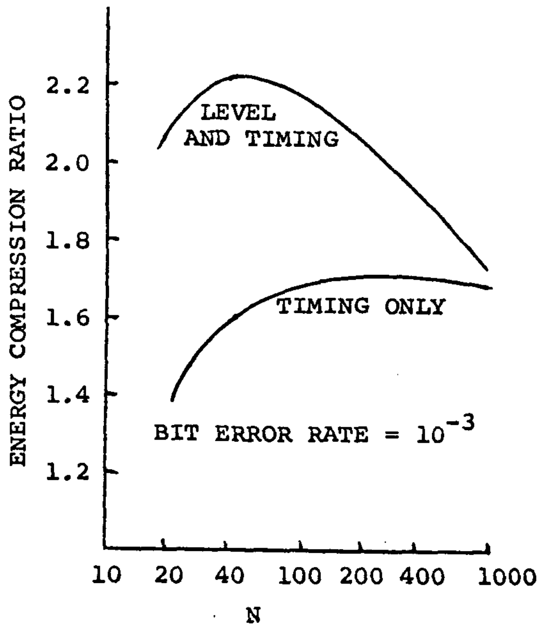

Fig. 4. ENERGY COMPRESSION RATIO USING (7,4) CODE

REFERENCES

1. L.D. Davisson, "The Concept of Energy Compression Ratio and Its Application to Run Length Coding," Final Report 1966 Session of the Goddard Summer Workshop, NASA Goddard Spaceflight Center Report No. X-700-67-94.

2. G.L. Raga, "Channel Noise-A Limiting Factor on the Performance of a Class of Adaptive Techniques," Conference on Adaptive Telemetry Held February 15-16, 1966, NASA Goddard Spaceflight Center Report No. X-731-67-7.

3. G.L. Raga, "Wideband Video Data Transmission," IEEE Trans. on Comm. Tech., Feb. 1967.